DTI ... COPY

(2)

ESD-TR-90-307

MTR-10226

AD-A223 304

Effects of Parameter Uncertainties on Software Development Effort Estimates

By

F. D. Powell

May 1990

Prepared for

Program Director, Communications and Airspace Management Systems Program Office
Electronic Systems Division
Air Force Systems Command
United States Air Force
Hanscom Air Force Base, Massachusetts

Approved for public release; distribution unlimited.

Project No. 027A

Prepared by

The MITRE Corporation
Bedford, Massachusetts

Contract No. F19628-86-C-0001

90 06 25 198

REVIEW AND APPROVAL

This technical report has been reviewed and is approved for publication.

Tracy A. Bombard

TRACY BOMBARD, 1LT, USAF
MMLSA Program Manager

David Kenyon

DAVID KENYON, GM-14
MLS Program Manager

FOR THE COMMANDER

David A. Herrell

DAVID HERRELKO, COL, USAF
Program Director
Comm. and Airspace Management SPO

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) MTR-10226 ESD-TR-90-307	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION The MITRE Corporation 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Burlington Road Bedford, MA 01730	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Program Director, (continued) 8b. OFFICE SYMBOL (If applicable) ESD/TGV	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F19628-86-C-0001
8c. ADDRESS (City, State, and ZIP Code) Electronic Systems Division, AFSC Hanscom AFB, MA 01731-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO: PROJECT NO: 027A TASK NO: WORK UNIT ACCESSION NO:

11. TITLE (Include Security Classification)
Effects of Parameter Uncertainties on Software Development Effort Estimates

12. PERSONAL AUTHOR(S)
Powell, Frederic D.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM ____ TO ____	1990 May	59

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	COCOMO Software Cost Estimates Software Effort

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The COnstructive COst MOdel (COCOMO) for software cost estimation uses inputs of size in lines of code and 15 Development Effort Multipliers (DEMs) that calibrate the estimate to the environment. This study examines the effects of uncertainties, expressed as probability densities of size and DEMs, on COCOMO effort estimates for software developments consisting of more than one Computer Software Configuration Item (CSCI). A Taylor series is used to find the mean and standard deviation of effort in terms of the means and standard deviations of the sizes and the DEMs of the CSCIs. The Central Limit Theorem then enables the evaluation of probability density and cumulative distribution functions for effort. Conditions under which this method is valid are demonstrated.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Judith Schultz	22b. TELEPHONE (Include Area Code) (617) 271-2844 22c. OFFICE SYMBOL Mail Stop D135

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

8a. Communications and Airspace Management Systems Program Office

EXECUTIVE SUMMARY

BACKGROUND

Software development estimating methods, such as the COnstructive COst MOdel (COCOMO) [1], provide a point estimate of the required effort in staff months. These estimating methods assume accurate a priori information for:

a. the size of the project in thousands of lines of code, and

b. the complexity of the project, the team's skill and familiarity with the language, the available equipment and tools, etc. These attributes are rated, and the estimated effort is adjusted accordingly by these Development Effort Multipliers (DEMs).

In a planning phase, there is significant uncertainty in these parameters. Consequently, a more appropriate and useful estimate is a probability cumulative distribution function (CDF) of effort.

PURPOSE

The purpose of this study is to compute the effort CDF and point estimate, given assumed probability density functions (pdfs) for the project's size and DEMs. In a planning phase, these estimates are frequently based on analogies to similar functions from other programs. Although there is considerable variability in the historical data, it is possible to specify realistic ranges for size and DEMs. Combining size and DEM uncertainty into a total system effort model is a difficult analytical problem and cannot be solved exactly, since the effort-estimating model is a nonlinear function of size, while the DEMs are multiplicative factors evaluated for the individual subsystems.

RESULTS

A procedure has been developed for estimating the system-level effort point estimate and CDF in terms of the means and variances of size and of the DEMs of the subsystems. Further, it is shown that the CDF of effort is determined to an accuracy which is adequate for software effort estimation whenever the largest contributing variance is not more than half the total variance of effort. The procedure is based on multivariable Taylor series and on the properties of the Central Limit Theorem.

CONCLUSIONS

The proposed solution is straightforward to implement, though there are several computational steps; a spreadsheet adequately supports the calculations. The method is recommended for use in planning estimates of software systems consisting of multiple functional areas, each with variability in the estimating parameters.

ACKNOWLEDGMENTS

This document has been prepared by The MITRE Corporation under Project No. 027A, Contract No. F19628-86-C-0001. The contract is sponsored by the Electronic Systems Division, Air Force Systems Command, United States Air Force, Hanscom Air Force Base, Massachusetts 01731-5000.

It is a pleasure to acknowledge the contributions of colleagues in D-93: Karen Pullen, Systems Analysis & Cost, Associate Department Head, who offered active support and encouragement, Michael Alexander and Paul Garvey whose conversations and suggestions were valuable, and Mrs. Dorothy Bertrand whose secretarial skill and patience are appreciated.

TABLE OF CONTENTS

TABLE OF CONTENTS (Concluded)

LIST OF ILLUSTRATIONS

LIST OF TABLES

SECTION 1

INTRODUCTION

1.1 BACKGROUND

Software development cost estimation is subject to a variety of uncertainties of which some are common to all systems, while others are unique to software. The requirements for any type of system may change in nature, scope, or functionality, or may vanish. Further, especially at the start of a software development process, there is considerable uncertainty as to the final investment of funds, effort, and schedule. One of the principal sources of this initial uncertainty is the estimated size of the software to be developed; the final size of a software system is sometimes more than twice the original estimate. Another major source of uncertainty lies in the estimating model of software development cost, for no highly accurate model exists at present. And it is common to find that the complexity, or other similar descriptors of a project's character, cannot be accurately predetermined. In spite of these important uncertainties, the project manager needs to know what budget has a desired probability of success, or what the probability of success is for a given budget. These are provided by the effort Cumulative Distribution Function (CDF).

A popular method for estimating software development cost is the COnstructive COst MOdel, COCOMO [1]. However, this is in principle only a method for forming point estimates for effort, staffing, and schedule, and has no inherent capability for estimating the effects of uncertainty of the size of the project and its various components, known as computer software configuration items (CSCIs). Further, this model requires selection of development effort multipliers (DEMs) to represent the impact of particular aspects of the project or environment, such as the complexity of the project, or the computation facility in which the development is conducted. Initially, these attributes also are inevitably uncertain due to incomplete knowledge of the product, the process, the facility, and the experience and ability of the staff. And, finally, COCOMO has a nontrivial error with respect to its own database. In consequence, there can be a significant range in the estimated cost of the system.

1.2 SCOPE

This report presents a method for extending COCOMO to incorporate the uncertainties of the size, the rating values selected for the DEMs, and the error of the model in determining the probability distribution of effort for a multiple CSCI project.

The reader is asssumed to be familiar with calculus and probability theory. It is not within the scope of this paper to comment on ESD acquisition and software cost estimation practices. The use of COCOMO is primarily intended to illustrate application of the method which is applicable to any software model of the aI^b form.

1.3 APPROACH

The analytical approach is outlined: a) the software development effort is expressed as a multivariable Taylor series in terms of the sizes and the DEMs of the CSCIs; b) the assumption that the uncertainties of these variables are statistically independent is used to approximate the mean effort and its standard deviation; c) the Central Limit Theorem (CLT) is then used to assert that the effort probability density function (pdf) and CDF are approximately normal.

1.4 DISCUSSION

The software cost analyst, when undertaking an uncertainty analysis, typically assigns a range of uncertainty to the estimated size of each major component. It is, however, very difficult for the analyst to state with any confidence whether the pdf is uniform, triangular, discrete, beta, or some other form. A similar observation applies to the DEMs. It is the particular virtue of the CLT that it shows the forms of the pdfs of the contributing elements to be unimportant: the forms of the pdf and CDF for the effort of the entire project converge to the familiar normal, or Gaussian, pdf and CDF as the number of CSCIs and DEMs increases. It is therefore legitimate for the analyst to assume very simple forms for the pdfs of size or DEMs. The CLT is thus especially appropriate and useful for software development effort uncertainty analysis.

1.5 OUTLINE OF REPORT

Section 2 presents the general approach to software development effort uncertainty analysis. Subsections 2.1 and 2.2 set forth, respectively, the foundations and the theoretical development of this approach. Various probability density functions considered suited to software DEMs and size estimation will be examined; their properties will be developed in section 3. The validity of the CLT in this application is considered in section 4. The analyst who is principally interested in the application and use of this technique, rather than its development, may go directly to section 5, where an example is fully developed. The various worksheet formats used in section 5 are gathered in the appendix.

SECTION 2

THE CENTRAL LIMIT THEOREM AND COCOMO

2.1 MATHEMATICAL FOUNDATIONS

In this section, analytical relationships will be developed from which the effort mean and standard deviation will be defined in terms of the means and standard deviations of size of the CSCIs and of the values of the DEMs. By invoking the CLT, the CDF for effort can be formed, which enables calculation of the probability that a specified value of effort will be exceeded, and the confidence level of the estimate can be determined. The approach is developed for Intermediate COCOMO. It can be extended to Detailed COCOMO, and to any model of the form aI^b.

To use the COCOMO software development cost estimating model requires the analyst to know the estimated size of each CSCI. In addition, the analyst requires the appropriate value to use for each of the 15 or more DEMs. These cost-driver attributes are not easily evaluated: for example, ACAP (analyst capability) is very low if "...the average analyst lies at the 15th percentile in terms of ability, efficiency, ability to communicate and cooperate." This is a highly subjective criterion. The other DEMs have similar definitions, with similar subjective evaluations. To specify a priori the true value for each of the DEMs is not possible. Thus, size and DEM rating levels are best defined by probability density functions.

The problem to be examined in this report is now stated: using the COCOMO equations, define the mean, the standard deviation, the pdf, and the CDF of effort for a multi-CSCI project in terms of the uncertainties of the CSCIs' sizes and DEMs, and the error of the effort estimating model.

The theoretical foundations of this approach are:

a. Law of Means: the mean of a sum of random variables taken from arbitrary distributions equals the sum of the means of the variables;

b. Law of Variances: the variance of a sum of independent random variables taken from arbitrary distributions equals the sum of the variances of the variables;

c. Taylor Series: a function which together with its derivatives is continuous in an interval may be exactly described everywhere within that interval in terms of the values of the variables, the function, and its derivatives, which are evaluated at any point within the interval;

d. CLT: the pdf and CDF of a sum of independent random variables converge to the normal as the variances of the contributors become small compared to the variance of the sum; the forms of the pdf and CDF of the sum do not depend on the forms of the pdf and CDF of the contributors.

The independent variables, assumed in this section, are:

a. The size of the software for each CSCI, measured in thousands of delivered source instructions (KDSI);

b. The rating levels of the DEMs chosen for each CSCI;

The dependent variable is development effort, measured in staff months (SM).

2.2 APPROACH

The approach is outlined:

a. An expression will be formed which defines effort in terms of the DEMs and the size in KDSI of the CSCIs;

b. Truncated Taylor series will be formed to define effort in terms of the independent variables;

c. The derivatives of the expressions in (a) will be formed and substituted into the Taylor series;

d. The mean and variance of effort will then be calculated under the assumption that the independent variables (size and DEMs) are random and statistically independent;

e. Finally, by invoking the CLT, it will be shown that the pdf and CDF of effort are normal under certain conditions, and this fact will be used to specify the probability that the effort estimate will not be exceeded.

The procedure outlined above is now set forth in detail.

2.2.1 Notation

D Development effort in project, SM

D_i Development effort in CSCI i, SM

N Nominal SM of effort in project

N_i Nominal SM of effort in CSCI i

I Number of KDSI in project

I_i Number of KDSI in CSCI i

σI_i Standard deviation of I_i

M_{ik} Value of DEM k in CSCI i

σM_{ik} Standard deviation of M_{ik}

M_i Effort adjustment factor, (product of all DEMs), in CSCI i

The independent random variables are I_i and M_{ik}; all other variables are dependent. The range of the index, i, is over all CSCIs, while the range of the index, k, is over all DEMs.

2.2.2 General Relationships

The effort is now defined in terms of the sizes of the CSCIs and the values of the DEMs, under the assumption that the uncertainties of the system's DEMs and size are defined at the level of the CSCIs rather than at lower levels. This assumption in no way limits the generality of the findings, but merely reduces the complexity of the notation and the derivations.

Following the COCOMO equations, the key relationships are:

Total size of project

$$I = \sum_i I_i \tag{2-1}$$

Computation of nominal SM

$$N = aI^b \tag{2-2}$$

where the parameters, a and b, are presented in table 2-1, see page 117 of [1].

Table 2-1. COCOMO Parameters

Development Mode	a	b
Organic	3.2	1.05
Semi-detached	3.0	1.12
Embedded	2.8	1.20

Computation of Effort Adjustment Factor for CSCI i, see page 125 of [1]:

$$M_i = \prod_k M_{ik} \tag{2-3}$$

Distribution of nominal SM among CSCIs

$$N_i = I_i N/I \tag{2-4}$$

The nominal effort is allocated to the several CSCIs in proportion to their sizes, as per instructions 4-6 on page 148 of [1]. The method used above in (2-4) for computing the distribution of nominal SM avoids the quantity nominal productivity used in [1], by substituting its definition.

Computation of development effort in CSCI i

$$D_i = N_i M_i \tag{2-5}$$

Computation of project total effort:

$$D = \sum_i D_i \tag{2-6}$$

Equations (2-1) through (2-6) completely define the foundations of Intermediate COCOMO to the extent required for the uncertainty analysis. As stated, effort as defined by D in (2-6) is the dependent variable, while the sizes of the CSCIs (I_i) in (2-1) and the DEMs' values given as M_{ik} in (2-3) are the independent variables. These independent variables are random and statistically independent.

2.2.3 Taylor Series

A Taylor series is formed for effort as a function of size, and of the DEMs, and is truncated after the second derivative terms. The error due to this truncation is demonstrated in section 4. The general form for this truncated multivariable Taylor series is

$$z = z_0 + \sum_i \frac{\partial z}{\partial x_i} (x_i - x_{i0}) + (1/2!) \sum_i \sum_j \frac{\partial^2 z}{\partial x_i \, \partial x_j} (x_i - x_{i0})(x_j - x_{j0}) \tag{2-7}$$

where z is the dependent variable, x_i and x_j are the independent variables and thus may be either size or DEMs, and the subscript 0 implies the multidimensional point about which the series is expanded. The indices range over all variables. In this case, the series is to be expanded about the mean values of the independent variables (I_i and M_{ik}), and takes the form

$$\begin{aligned} D = D_0 &+ \sum_i \frac{\partial D}{\partial I_i} (I_i - \bar{I}_i) + \sum_i \sum_k \frac{\partial D}{\partial M_{ik}} (M_{ik} - \bar{M}_{ik}) \\ &+ (1/2!) \sum_i \sum_j \frac{\partial^2 D}{\partial I_i \, \partial I_j} (I_i - \bar{I}_i)(I_j - \bar{I}_j) \\ &+ (1/2!) \sum_i \sum_j \sum_k \frac{\partial^2 D}{\partial I_i \, \partial M_{jk}} (I_i - \bar{I}_i)(M_{jk} - \bar{M}_{jk}) \\ &+ (1/2!) \sum_i \sum_j \sum_k \frac{\partial^2 D}{\partial M_{ik} \, \partial M_{jk}} (M_{ik} - \bar{M}_{ik})(M_{jk} - \bar{M}_{jk}) + \ldots \end{aligned} \tag{2-8}$$

D_0 is the usual point estimate of the development effort based on mean sizes and mean values of the DEMs, in accordance with (2-1) - (2-6). The partial derivatives are to be evaluated at the mean values of size and DEMs. The overbar implies the mean values for the sizes, I_i or I_j, and for the DEMs, M_{ik} or M_{jk}.

We now may formulate expressions for the mean value of D and its variance, and evaluate (2-8) under the following assumptions:

a. The uncertainty of size of any CSCI is independent of the uncertainty of size of the other CSCIs;

b. The uncertainty of size of any CSCI is independent of the uncertainties of the DEMs for all CSCIs;

c. The uncertainty of the value of any DEM is independent of the uncertainties of all other DEMS so that all the random variables are mutually statistically independent.

The mean value of (2-8) is determined. The three terms in (2-7) expand into the six terms of (2-8) due to the presence of variables of two types (I_i and M_{ik}); it will be shown that the assumptions noted above cause four terms to vanish, and the procedure for evaluating the other two will be demonstrated. Consider the first term on the right of (2-8); as stated above, this is the development effort point estimate determined by (2-1) through (2-6) evaluated with the mean values of size and DEMs.

Now consider the second and third terms, which appear on the first line of (2-8) and are the first derivative terms. The expected value of $(I_i-\bar{I}_i)$ is zero since the expected value of I_i is $\bar{I}_i$; the expansion is about the mean of the variables, $\bar{I}_i$, therefore the second term vanishes. By the same reasoning, the expected value of $(M_{ik}-\bar{M}_{ik})$ is zero, and that term vanishes also.

The fourth term on the right of (2-8), in the second line, is reduced by the expectation operator and assumption (1), above, to the mean value of

$$(1/2) \sum_i \frac{\partial^2 D}{\partial I_i{}^2}(I_i-\bar{I}_i)^2, \text{ which is } (1/2) \sum_i \frac{\partial^2 D}{\partial I_i{}^2} \sigma^2 I_i$$

since the expected value of the product is zero unless i=j.

Assumption (2) causes the fifth term, on the third line of (2-8), to vanish in the presence of the expectation operator.

The sixth term on the right of (2-8) is evaluated by using assumption (3) and the discussion for the fourth term, and yields

$$(1/2) \sum_i \sum_k \frac{\partial^2 D}{\partial M_{ik}{}^2} \sigma^2 M_{ik}$$

which vanishes as the second and higher partial derivatives of effort with respect to M_{ik} are identically zero. Third and higher partial derivatives of effort with respect to I_i are neglected; the consequent errors will be discussed in section 4.

Combining the results developed above yields the estimate of the mean effort

$$\bar{D} = D_0 + (1/2) \sum_i \frac{\partial^2 D}{\partial I_i{}^2} \sigma^2 I_i \tag{2-9}$$

The second term of (2-9) is relatively small. Its value will be demonstrated in the example shown in section 4.2.2; see also column 10 of figure 5-6.

The variance of effort is now determined. The assumptions of mutual independence of the random variables will be used in conjunction with the properties of the expectation operator. Subtract (2-9) from (2-8); this yields an expression for the uncertainty of effort about the mean. Then square and take the expectation; these steps yield the variance of effort as

$$\sigma^2 D = \sum_i \left(\frac{\partial D}{\partial I_i} \right)^2 \sigma^2 I_i + \sum_i \sum_k \left(\frac{\partial D}{\partial M_{ik}} \right)^2 \sigma^2 M_{ik} \qquad (2\text{-}10)$$

In principle, the variance must also depend on all the odd higher derivatives with respect to I_i and M_{ik}. But these are identically zero for M_{ik}, while the derivatives with respect to I_i are very small. The examples in section 4 will show the error of this approach for simple cases.

2.2.4 Derivatives

The following partial derivatives are required:

$$\frac{\partial D}{\partial I_i}, \quad \frac{\partial D}{\partial M_{ik}}, \quad \text{and} \quad \frac{\partial^2 D}{\partial I_i^{\,2}}$$

Consider first the partial derivative with respect to the size of CSCI i. Relationships (2-1) - (2-4) enable formation of this partial derivative, which is a size-sensitivity coefficient, as

$$\frac{\partial D}{\partial I_i} = [\ (b-1)D_0 + \bar{N}\bar{M}_i\ ]/\ \bar{I} \qquad (2\text{-}11)$$

In forming this derivative it is essential to remember that I_i is included in I, the total size of the project, as shown in (2-1).

Partial differentiation of (2-11) with respect to I_i, and use of (2-1) through (2-6) and (2-11), yields the required second derivative of effort with respect to size

$$\frac{\partial^2 D}{\partial I_i^{\,2}} = (b-1)[(b-2)D_0 + 2\bar{N}\bar{M}_i]/\bar{I}^2 \qquad (2\text{-}12)$$

Similarly, the derivative of effort with respect to a DEM is

$$\frac{\partial D}{\partial M_{ik}} = \bar{N}_i \prod_{j \neq k} \bar{M}_{ij} = \bar{D}_i / \bar{M}_{ik} \qquad (2\text{-}13)$$

As stated after (2-7), the partial derivatives must be evaluated with mean sizes and mean values of the DEMs.

Substitution of (2-12) into (2-9), and of (2-11) and (2-13) into (2-10), yields the estimated mean development effort as

$$\bar{D} = D_0 + (1/2) \sum_i \{(b-1)*[(b-2)D_0 + 2\bar{N}\bar{M}_i]/\bar{I}^2\} \sigma^2 I_i \qquad (2\text{-}14)$$

and the variance of development effort as

$$\sigma^2 D = \sum_i \{[(b-1)D_0 + \bar{N}*\bar{M}_i]/\bar{I}\}^2 \sigma^2 I_i + \sum_i \sum_k (\bar{D}_i/\bar{M}_{ik})^2 \sigma^2 M_{ik} \qquad (2\text{-}15)$$

The mean effort and the variance of effort have thus been defined in terms of quantities which are, or can be made to be, readily available in a computational format for Intermediate COCOMO.

2.2.5 Cumulative Distribution Function of Effort

The procedures described above enable formation of the variance and mean value of the estimated SM of effort in terms of the means and variances of size of the CSCIs and of the DEMs. In addition to the uncertainties of the size and DEMS, the error of the COCOMO model may be included in the determination of the variance of the estimate. The CLT justifies disregarding whether the model's errors are exactly normally distributed. The errors of the COCOMO model are independent of the errors due to uncertainties of size or of DEMs, and therefore the variance of the model's estimate of effort may be added to those due to size and DEMS to present an overall assessment of the variance of effort. The CLT is now used to assert that, as the number of independent random variables increases, the pdf and CDF of their sum converge to the normal pdf and CDF, which are fully determined by the mean and variance of effort. Any packaged program or standard table of the normal probability integral may be used to calculate the CDF; see also the appendix.

SECTION 3

PROBABILITY DENSITY FUNCTIONS FOR PARAMETER UNCERTAINTIES

It was assumed in section 2 that the mean and standard deviation for pdfs appropriate for software risk analysis are available. In the present section, these properties will be derived for uniform, triangular, and discrete pdfs. Discrete pdfs are especially appropriat. for characterizing the uncertainties of DEMs. It is considered that more sophisticated pdfs such as beta are inappropriate for software analysis as they require the analyst to have great insight into the nature of the uncertainties of size or of the DEMs. In reality, it is difficult to select reasonable upper and lower bounds for a uniform distribution, which has only two parameters! Moreover, the effect of the CLT elimin^tes the significance of fine distinctions in the forms of assumed pdfs. The validity of the very simple geometric forms for analysis using the CLT will be demonstrated in section 5.

The use of Taylor series in section 2 requires that the independent variables be continuous. This is not strictly true for size, since the size of a piece of code is defined in lines of code, which is an integer quantity; it is hard to conceive of a fraction of a line of code. Nonetheless, as the minimum increment of one line is 0.001 KDSI, it is reasonable to assume that effort is a continuous function of the continuous variable of size. The DEMs are fundamentally continuous variables; however, they are known only at the specific tabulated points and therefore a discrete pdf structure is useful.

3.1 UNIFORM DISTRIBUTION

A uniform probability density function has the definition

$$f(x) = \begin{cases} 0, & x < L \\ 1/(H-L), & L \leq x \leq H \\ 0, & H < x \end{cases} \tag{3-1}$$

where L Minimum value of x
H Maximum value of x
x Statistical variable

The mean of this pdf is easily found from the geometry to be

$$\bar{x} = (H-L)/2 + L = (H+L)/2 \tag{3-2}$$

The variance is

$$\sigma^2 = \int_L^H x^2 f(x)dx - (\bar{x})^2 = (H-L)^2/12 \qquad (3\text{-}3)$$

3.2 TRIANGULAR DISTRIBUTION

The mean and variance of a general triangular pdf are derived. Figure 3-1 shows a triangular pdf and its notation.

$$f_- = [2/(H-L)][(x-L)/(M-L)]$$

$$f_+ = [2/(H-L)][(H-x)/(H-M)] \qquad (3\text{-}4)$$

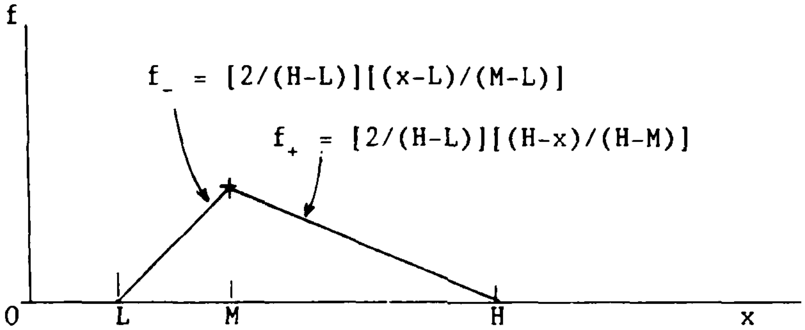

Figure 3-1. Triangular pdf and Notation

Notation

- f Probability density
- L Minimum value of x
- M Most probable value of x, mode
- H Maximum value of x
- x Statistical variable, thousands of lines of code

The mean is derived according to the rule,

$$\bar{x} = \int_{-\infty}^{\infty} x\ f(x)\ dx = \int_L^M x\ f_-(x)\ dx + \int_M^H x\ f_+(x)\ dx = (L+M+H)/3 \qquad (3\text{-}5)$$

where $\bar{x}$ is the mean.

The variance is calculated according to the rule

$$\sigma^2 = \int_{-\infty}^{\infty}(x-\bar{x})^2 f(x)\, dx = \int_{L}^{M}(x-\bar{x})^2 f_{-}(x)\, dx + \int_{M}^{H}(x-\bar{x})^2 f_{+}(x)\, dx$$

$$= \{[(L^2+M^2+H^2)-(LM+LH+MH)]/18\} \tag{3-6}$$

A right-triangle pdf is sometimes used in software size analysis. The mean and variance for this special form are, if L=M is assumed,

$$\bar{x} = (2L+H)/3 \tag{3-7}$$

and

$$\sigma^2 = (H-L)^2/18 \tag{3-8}$$

3.3 DISCRETE DISTRIBUTION

A discrete distribution is occasionally realistic for size and is usually ideal for DEMs, as they are given only for specific values in [1]. For example: the probability of DEM rating 1 is P_1, of rating 2 is P_2, of rating 3 is P_3, etc., where the sum of the P_k must equal unity. Then, defining

x_k The rating values tabulated in [1] for a DEM
P_k The probability of those rating values for that DEM

the mean value of the DEM is

$$\bar{x} = \sum_k x_k P_k \tag{3-9}$$

and the variance of the DEM is

$$\sigma^2 = \sum_k (x_k-\bar{x})^2 P_k \tag{3-10}$$

SECTION 4

ACCURACY AND VALIDITY

This section is devoted to resolving two important questions:

a. Are the means and variances of effort determined with accuracy sufficient to be confidently used?

b. Under what conditions is the CDF derived by this procedure sufficiently accurate?

These questions are now considered in the order set forth above.

4.1 ACCURACY OF MEANS AND VARIANCES

The accuracy with which the Taylor series method determines the mean and variance of a multi-CSCI software system is considered. The approach is outlined. There exists a method, elegantly presented in [4] for uniform and triangular pdfs of size, which enables exact calculation of the mean and variance of effort for a single CSCI. The mean and variance of a single CSCI determined by the Taylor series method will therefore be compared to the exact values. The following rules will then be used to infer that the results may be extended to the multi-CSCI case:

a. The mean of a sum of random variables equals the sum of the means of those variables;

b. The variance of a sum of independent random variables equals the sum of the variances of the variables.

These rules are valid in general.

The mean and standard deviation of several uniform pdfs are calculated by the Taylor series and exact methods and are compared in table 4-1, where the COCOMO Embedded Mode values a=2.8 and b=1.2 have been used; this choice of the parameter, b, exhibits the maximum error for the Taylor series method presented in this report. The errors of the mean and standard deviation are due to the truncation of the Taylor series.

Table 4-1. Mean and Standard Deviation of Effort for Several Uniform pdfs of Size

Size-Range Low/High KDSI	Mean Effort Taylor Series	Mean Effort Exact Method	Effort Standard Deviation Taylor Series	Effort Standard Deviation Exact Method
4/16	45.02	45.03	18.4	18.3
16/64	237.6	237.7	97.4	96.8
64/256	1254	1255	514	511
256/1024	6619	6622	2712	2697

The largest error in the values of the means is less than one part per thousand, while the largest error in the standard deviations is less than six parts per thousand. These errors are considered entirely satisfactory, as they are several orders of magnitude smaller than the errors of the COCOMO model. These errors are due to the truncation of the Taylor series.

Similar results, shown in table 4-2, are found for a group of various triangular pdfs of size. The mean and standard deviation of effort for the exact and Taylor series methods again compare closely. The differences are small compared to the COCOMO model error, with standard deviation of approximately 15 to 20 percent of the point estimate of effort.

Table 4-2. Mean and Standard Deviation of Effort for Triangular pdfs of Size

Size-Range Low/Mode/High KDSI	Mean Effort Taylor Series	Mean Effort Exact Method	Standard Deviation of Effort Taylor Series	Standard Deviation of Effort Exact Method
16/16/64	181.9	181.8	76.0	77.1
16/24/64	199.4	199.4	71.7	72.4
16/32/64	217.5	217.5	69.2	69.4
16/40/64	235.91	235.94	68.8	68.6

The combined effects of skew and truncation of the Taylor series appear in the first line of this table, where the standard deviations differ by 1.4 percent. The bottom row shows the effects of truncation only. The results presented in tables 4-1 and 4-2 show that the Taylor series method enables calculation of the mean and standard deviation (or variance) of effort to an acceptable degree of accuracy for uniform or triangular pdfs.

4.2 CONDITIONS FOR CALCULATING THE CDF

It will now be demonstrated that the CDF of a multi-CSCI system can be determined to a sufficient degree of accuracy by combining the Taylor series method with the CLT. The CLT shows that the CDF of a sum of independent random variables converges to the normal CDF as the number of such variables increases, without regard to the forms of the contributing pdfs. The approach which will be used to determine the conditions under which the CLT may be used in the present context is to show that under various worst-case conditions the following conjecture is true:

> The CDF of effort may be calculated under the assumption that it is normal whenever the largest contributing variance of size or of the DEMs does not exceed half the total variance of the effort.

As stated, the contributing variances are those of size and of the DEMs in the various CSCIs; the error of the model is excluded. This condition is viewed as sufficient but not necessary. The conjecture will be supported by examples which satisfy the condition cited above. These examples are:

> The variables come from two identical, independent, pdfs of size which are either uniform or right-triangular. The DEMs are assumed to be nominal, and the error of the COCOMO model is neglected.

In these examples, the two contributing variances are equal, and the condition for the conjecture is marginally satisfied. Neglecting the approximately normal error of the model is highly conservative.

The criterion for acceptance of the Taylor series model of CDF is stated: this model will be considered acceptably accurate if its worst error of estimated effort is less than 5 percent of the point estimate of effort. In one particular area of interest, the range from 60 to 90 percentile, the Taylor series model is accurate within 2 percent or less, an error which can be considered negligible for cost-estimating purposes in the planning phase.

4.2.1 A Sum from Two Uniform pdfs

Assume two identical uniform pdfs, f_1 and f_2, defined by

$$f_i = \begin{cases} 0, & x < L_i \\ 1/(H_i - L_i), & L_i \le x \le H_i \\ 0, & H_i < x \end{cases} \qquad (4\text{-}1)$$

where i = 1, 2. Also assume these pdfs are identical so that $L_1 = L_2 = L$, and $H_1 = H_2 = H$, and $R = H_1 - L_1 = H_2 - L_2 = H - L$. The pdf of the sum (S) of the variates is, see [3], the convolution integral

$$f(S) = \int f_1(x)\, f_2(S-x)\, dx \qquad (4\text{-}2)$$

Using the definitions of f_1 and f_2 and (4-3) yields the pdf of the sum, S, as

$$f(S) = \begin{cases} 0, & S < L_1+L_2 \\ \int_{L_1+L_2}^{S} dx/R^2 = [S-(L_1+L_2)]/R^2, & L_1+L_2 \le S < L_1+L_2+R \\ \int_{S}^{H_1+H_2} dx/R^2 = [H_1+H_2-S]/R^2, & L_1+L_2+R \le S \le H_1+H_2 \\ 0, & H_1+H_2 < S \end{cases} \qquad (4\text{-}3)$$

This triangular pdf is symmetrical about the value $S=L_1+L_2+R = H+L$.

Assume that this symmetrical triangular pdf of size has Low=16 KDSI, Mode=40 KDSI, and High=64 KDSI, and use the COCOMO parameters a=2.8, and b=1.2. The mean and standard deviation of effort for this case are presented in table 4-2. These estimated values are now used to form the CDF, using the assumption that the distribution is normal, consistent with use of the CLT. This normal CDF is compared in table 4-3, below, with the exact CDF for the symmetrical triangular pdf defined by (4-3), calculated by the method of [4].

The difference of greatest magnitude (*) is 0.017, which corresponds to an error of effort of 5 SM. This error of effort is negligibly small compared to the selected criterion.

Uniform pdfs are symmetrical, and the pdf of the sum of two identical uniform pdfs is a symmetrical triangle. It is not surprising that such a sum has nearly normal distribution. Right-triangular pdfs are, however, at the extreme of asymmetry. This case is now examined.

Table 4-3. Normal CDF, True CDF, and Difference for a Symmetrical Triangular pdf, 16K to 64K Size

Effort SM	Normal CDF	True CDF	Difference
70	0.008	0.000	0.008
78.001(Low)	0.011	0.000	0.011
100	0.024	0.012	0.012
125	0.054	0.051	0.003
150	0.106	0.117	- 0.011
175	0.188	0.205	- 0.017 *
200	0.301	0.316	- 0.015
235.908(Mean)	0.500	0.510	- 0.010
250	0.581	0.589	- 0.008
275	0.715	0.710	0.005
300	0.824	0.809	0.015
325	0.902	0.886	0.016
350	0.951	0.943	0.008
375	0.978	0.980	- 0.002
400	0.991	0.998	- 0.007
411.6935(High)	0.995	1.000	- 0.005
425	0.997	1.000	- 0.003

4.2.2 A Sum from Two Right-Triangular pdfs

Asssume two identical right-triangular pdfs, defined without loss of generality on the interval $0 \leq x \leq 1$, as shown in figure 4-1 below.

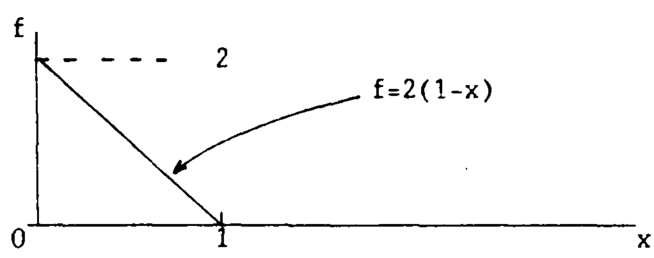

Figure 4-1. A Right-Triangular Probability Density Function

The pdf of the sum of the two variates is

$$f(S) = \begin{cases} 0, & S < 0 \\ 4\int_0^S (1-x)(1-S+x)dx = 2S(6-6S+S^2)/3, & 0 \leq S < 1 \\ 4\int_{S-1}^1 (1-x)(1-S+x)dx = 2(8-12S+6S^2-S^3)/3, & 1 \leq S \leq 2 \\ 0, & 2 < S \end{cases} \qquad (4\text{-}4)$$

As before, integration of (4-4) yields the CDF of the sum as

$$F(S) = \begin{cases} 0, & 0 < S \\ S^2(S^2-8S+12)/6, & 0 \leq S < 1 \\ (32S-24S^2+8S^3-S^4-10)/6, & 1 \leq S \leq 2 \\ 1, & 2 < S \end{cases} \qquad (4\text{-}5)$$

Following the procedure of [4], (4-5) may be transformed to the required form by the substitutions

$$S=[t-(L_1+L_2)]/R,\ R=H-L,\ \text{and}\ (D/\alpha)^{1/b}=t \qquad (4\text{-}6)$$

where, using the notation of section 2, $\alpha = a(\sum_i \bar{I}_i \prod_j \bar{M}_{ij})/\bar{I}$ (4-7)

The CDF of the sum is calculated as follows:

a. Calculate α from (4-7) and then select a value for D;
b. Using (4-6), calculate t and then S;
c. Given S, calculate the CDF from (4-5).

The cumulative probability distribution thus generated is compared to the normal CDF in table 4-4. The mean and variance used for calculating the normal CDF are calculated here.

The size data for the assumed right-triangular pdfs are:

Low=L_1=L_2=16 KDSI, High=H_1=H_2=64 KDSI.

Mean sizes of the pdfs are, from (4-4):

$$\bar{I}_1=\bar{I}_2=(16+16+64)/3=32 \text{ KDSI}.$$

Size standard deviations are, from (4-6):

$$\sigma I_1=\sigma I_2= \{[(16^2+16^2+64^2)-(16*16+16*64+16*64)]/18\}^{1/2} = 8\sqrt{2}.$$

The DEMs have been assumed nominal and without uncertainty so that their products are unity and their variances are zero; then the point estimate $D_0=\bar{N}$ as a consequence. The point estimate of effort is

$$D_0=2.8(32+32)^{1.2}=411.693 \text{ SM}$$

The first and second derivatives with respect to size are, from (2-11)

$$\frac{\partial D}{\partial I_1} = \frac{\partial D}{\partial I_2} = \{[(1.2-1)411.693+411.693]/64\}=7.719 \text{ SM/KDSI}$$

and, from (2-12)

$$\frac{\partial^2 D}{\partial I_1{}^2}= \frac{\partial^2 D}{\partial I_2{}^2} = \frac{(1.2-1)[(1.2-2)411.693+2*411.693]}{(64*64)} = 0.024$$

The estimated mean effort is, from (2-14)

$$\bar{D} = 411.69+0.024[(8\sqrt{2})^2+(8\sqrt{2})^2]/2=411.69+3.07=414.8$$

The <u>exact</u> mean effort is 421.0 SM, from [4].

The estimated standard deviation of effort is, from (2-15),

$$\sigma_D = \frac{[(1.2-1)411.69+411.69]}{64} [(8\sqrt{2})^2+(8\sqrt{2})^2]^{1/2} = 123.5$$

The <u>exact</u> standard deviation is 125.1, from [4].

The estimated mean and standard deviation of effort are used to calculate the Taylor series model's CDF, which is normal under the CLT hypothesis; this CDF is presented in table 4-4, below. The exact CDF is also presented, together with the error of probability between the two CDFs at the same effort.

Table 4-4. Exact and Normal Cumulative Probability Distributions of Effort for the Sum of Variates from Two Right-Triangular pdfs

No. of Std. Devs.	Effort	Normal CDF	Exact CDF	Difference
- 1.98	170	0.0237	0	0.0237
- 1.91	179.2	0.0282	0.0000	0.0282
- 1.74	200	0.0410	0.0078	0.0332
- 1.54	225	0.0622	0.0352	0.0270
- 1.33	250	0.0911	0.0783	0.0128
- 1.13	275	0.1289	0.1334	-0.0045
- 0.93	300	0.1764	0.1975	-0.0211
- 0.73	325	0.2336	0.2676	-0.0240
- 0.52	350	0.2999	0.3415	-0.0416
- 0.32	375	0.3737	0.4170	-0.0433 *
- 0.12	400	0.4524	0.4922	-0.0398
0.083	425	0.5330	0.5653	-0.0323
0.29	450	0.6122	0.6350	-0.0228
0.49	475	0.6871	0.6997	-0.0126
0.69	500	0.7549	0.7582	0.0033
0.89	525	0.8139	0.8096	0.0043
1.09	550	0.8632	0.8528	0.0104
1.30	575	0.9027	0.8881	0.0146
1.50	600	0.9331	0.9163	0.0168
1.70	625	0.9556	0.9388	0.0168
1.90	650	0.9716	0.9562	0.0154

The difference is relatively small everywhere except in the region from 0.52 to 0.12 standard deviations below the mean effort; the difference of greatest magnitude is marked by a *. This region is of relatively little interest, for the principal concern of a project manager is usually in the over-run range of cumulative probabilities from 70 to 90 percent.

The differences, in column 5 of table 4-4, are the errors of the cumulative probability distribution function as a function of effort. These errors may be converted to errors of the estimate of effort as a function of the cumulative probability; this structure of errors is presented in table 4-5, below.

Table 4-5. Errors of Estimated Effort

Probability	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Error of Effort	+6	-10	-15	-13	-11	-8	-4	-2	+9

The maximum error of effort, in table 4-5, is -15 SM; this error is less than five percent of the point estimate (411 SM), and is also less than 1/3 of the approximately 62 SM error of the COCOMO model. Further, the errors in the region of interest from 50 to 90 percent probability are even smaller. Therefore, even in this extreme case the error of the Taylor series model is acceptably small compared to the selected criterion.

The accuracy of this method improves as the number of error sources increases.

4.3 INFERENCES

It has been shown that the sum of variables from either two uniform, or two right-triangular, identical pdfs yields a CDF which is sufficiently close to normal to be accepted as such, under the selection criterion proposed above. It will now be shown that the CLT may be used to extend the examples to the more general situation of more than two contributing variances.

Assume that the variance of the system is comprised of two large and equal variances, plus one or more smaller contributors. The effect of these smaller contributors is to make the pdf and CDF of the overall system closer to normal, by virtue of the action of the CLT. Alternately, assume that instead of one or more small contributors there is only the error of the COCOMO model itself. But this error is, on inspection, nearly normal, taking into account that there are only 63 cases in the COCOMO database, and therefore the CDF of the sum of the two large contributors and the error of the model is more nearly normally distributed than merely the sum of the two principal contributors.

The sum of variables from two general triangles with the same size-range cannot be further from normal than the sum from two identical right-triangular pdfs. Therefore, the example considered above is a worst case; any other combination of triangles must be more nearly normal.

It is therefore concluded that the conjecture proposed at the beginning of this section is demonstrated. The conclusions are:

a. The Taylor series method may legitimately be used in all cases to determine the mean and standard deviation of effort;

b. The Taylor series/CLT method may be used to find the cumulative distribution of effort whenever the largest contributing variance due to size or to a DEM does not exceed half of the total variance of effort.

It should be observed that the criterion stated above is not valid in general. However, it is at present valid in software effort analysis because the error of software effort estimating models is large. If a software effort estimating model should appear with much smaller errors, then this method of effort uncertainty analysis, and the validity criterion which was offered, must be reconsidered.

P. R. Garvey asserts that his method [4] for calculating the effects of uncertainties of size can be extended from the single-CSCI case, for which it is exact, to the multiple-CSCI case by use of the CLT. The analysis, above, in this section, implies that validity of his assertion requires satisfying the same criteria used here for Taylor series. His assertion is therefore justified, and his approach may be used for the size aspects of a multiple-CSCI case.

4.4 SUGGESTIONS FOR THE SOFTWARE ANALYST

Some suggestions for the software cost analyst are offered.

Size: The uncertainty of size of a software project is primarily at the CSCI level, and is related to the number, rather than the size, of modules at the lowest level of the hierarchy of components. Triangular or uniform pdfs are reasonable ways to describe size uncertainty.

DEMs: At least some DEMs may reasonably be assumed to have uncertainties. The structure of the algorithms which have been developed enables setting the uncertainty of a DEM directly as, for example, "standard deviation for DEM 36 = 3 percent, i.e., 0.03." This follows from the definition of DEMs as multipliers with nominal value of unity; see also the further discussion of this point in section 5. The analyst may therefore express an opinion of the uncertainty of selection of the DEM as a fraction of the DEM's value, or may use the discrete pdf approach outlined in section 4.3; the latter is recommended.

SECTION 5

EXAMPLE

An example of the computation procedure is presented in this section. A system comprised of three CSCIs is assumed. The sizes of the CSCIs, and the values of three DEMs in each CSCI, are assumed to be uncertain. The mean effort and the standard deviation of effort will be evaluated, and the normal probability cumulative distribution function which results will be determined.

The procedure is conducted by using work sheets which are specifically designed for Intermediate COCOMO.

The computation work sheets are of two distinct types. The first type is used to compute the mean values of the DEMs, their standard deviations, and the other related quantities. One of these forms is filled out for each CSCI. The second form is used to calculate the mean and standard deviation of effort, including the effects of size and of DEMs; this form is filled out once for the project.

5.1 DATA

The data for the example are gathered here. Table 5-1 presents the assumed uncertainties of size for the three CSCIs.

Table 5-1. Assumed Size Uncertainties

CSCI #	Type of Distribution	Low-Size L	Mode-Size M	High-Size H
1	Triangular	16	32	64
2	Right-Triangular	40	40	80
3	Uniform	40	-	80

No value is entered in table 5-1 for the mode for CSCI 3 as its pdf is assumed to be uniform. The data from table 5-1 will be used in figure 5-6.

Table 5-2 presents the assumed uncertainties for three DEMs in each of the three CSCIs. DEMs not entered in table 5-2 are assumed to be nominal, with value unity and no uncertainty. Consistent with the prior discussions, the DEMs' uncertainties are described by discrete distributions.

Table 5-2. Assumed Uncertainties of DEMs

	CSCI	DEM Probabilities					
#	Name	Very Low	Low	Nom	High	Very High	Extra High
1	CPLX	0.05	0.10	0.25	0.35	0.20	0.05
	TIME			0.50	0.50		
	ACAP		0.10	0.30	0.50	0.10	
2	DATA			0.70	0.30		
	STOR				0.60	0.40	
	PCAP		0.50	0.50			
3	RELY				0.50	0.50	
	AEXP		0.25	0.50	0.25		
	LEXP		0.20	0.70	0.10		

These data will be used in figures 5-2 through 5-4.

5.2 REORGANIZATION OF MEAN AND VARIANCE FORMULATIONS

It is useful, for computational purposes, to reorganize the equations for mean effort and its standard deviation. The modified forms are presented here; it will be observed that the changes are:

Size: The first partial derivative with respect to size is premultiplied by mean size, and the standard deviation of size is divided by the mean size; with the second partials, the square of size is used.

DEM: The partial derivative with respect to a DEM is multiplied by that DEM's mean value, and the standard deviation of the DEM is divided thereby.

These changes are thus only of form. The equations for mean effort and variance of effort become

Mean effort:

$$\bar{D} = D_0 + \sum_i (1/2)\{(b-1)[(b-2)D_0 + 2\bar{N}\bar{M}_i]\}^2(\sigma I_i/\bar{I})^2 \tag{5-1}$$

Variance of effort:

$$\sigma^2 D = \sum_i [(b-1)D_0+\bar{N}\bar{M}_i]^2(\sigma I_i/\bar{I})^2 + \sum_i \sum_j \bar{D}_i{}^2(\sigma M_{ij}/\bar{M}_{ij})^2 \tag{5-2}$$

The inner summation of the second term of (5-2) may be simplified as

$$(\sigma M_i/\bar{M}_i)^2 = \sum_j (\sigma Mi_j/\bar{M}_{ij})^2 \tag{5-3}$$

The notation is simplified by defining

$$S_i = (b-1)D_0+\bar{N}\bar{M}_i \tag{5-4}$$

$$T_i = (b-1)[(b-2)D_0 + 2\bar{N}\bar{M}_i] \tag{5-5}$$

The equations for calculating the mean and variances of uniform, triangular, and discrete pdfs are gathered here for convenience.

Means

Uniform pdf $\bar{x} = (L + H)/2$ (5-6)

Triangular pdf $\bar{x} = (L + M + H)/3$ (5-7)

Discrete pdf $\bar{x} = \sum_k x_k P_k$ (5-8)

Variances

Uniform pdf $\sigma^2 = (H - L)^2/12$ (5-9)

Triangular pdf $\sigma^2 = [(L^2+M^2+H^2)-(LM+LH+MH)]/18$ (5-10)

Discrete pdf $\sigma^2 = \sum_k(x_k-\bar{x}_k)^2P_k$ (5-11)

where x_k are the values of the ratings of a DEM, and P_k are the probabilities of these ratings.

5.3 MEAN AND STANDARD DEVIATION OF DEMs

Computation of the mean, standard deviation, and CDF of effort requires computation of the means and standard deviations of the DEMs. Figures 5-1 through 5-4 present the computation of the mean and standard deviation of effort due to the uncertainty of the DEMs. Figure 5-1 is a detailed list of instructions, while figures 5-2, 5-3, and 5-4 present the actual calculations and data for the DEMs of the three CSCIs, consistent with the data in table 5-2 and the second term in (5-2), or the reformulation in (5-3).

The output from this stage of the computation appears at the bottom of columns 13, 14, and 15 on figures 5-2 through 5-4. These outputs will be used in computing the standard deviation of effort in figures 5-5 and 5-6, which will be discussed next.

5.4 MEAN AND STANDARD DEVIATION OF EFFORT

Computation of the mean and standard deviation of effort, and its CDF, requires computation of the means and standard deviations of size of the CSCIs. Figure 5-5 is the detailed instruction set for using figure 5-6. Figure 5-6 is used for calculating the point estimate of effort, the estimated mean effort, and the standard deviation of effort. The size data from table 5-1 are entered into columns 1, 2, and 3. The result of the computations on the DEMs, from figures 5-2, 5-3, and 5-4, are entered into column 6. Figure 5-5 presents the procedure for using this form. Any standard table of the normal probability function may be used to find the CDF; alternately, see a procedure in [5].

5.5 CUMULATIVE DISTRIBUTION FUNCTION OF EFFORT

Given the mean effort, its standard deviation, and the assumption that the distribution is normal, the CDF is easily calculated by using a standard table of the normal probability function. This table is entered with x, the number of standard deviations between any value of effort, D', and the mean effort, and is defined as

$$x = (D' - \bar{D})/\sigma_D \qquad (5\text{-}12)$$

Table 5-3 presents the probability cumulative distribution function, CDF, which follows from the results of the computations in figure 5-6; that figure shows the mean effort to be 1412 SM, and the standard deviation of effort to be 322.9 SM.

This form is used for calculating the mean and the standard deviation of the product of the DEMs in each CSCI for Intermediate COCOMO. The outputs are the mean product of the DEMs and the normalized standard deviation of the product of the DEMs. The mean appears at the bottom of column 13, while the normalized standard deviation appears at the bottom of col. 15, and the number of uncertain items is given at the bottom of col. 14. Detailed instructions follow.

1. At upper-right corner, enter the Mode (Organic, Semi-det...., Embedded), and the COCOMO coefficients a and b for that mode. Inside form, at upper-left corner, enter the CSCI number (i=).
2. The *ratings* for each attribute for Intermediate COCOMO are given in columns 1, 3, 5, 7, 9, 11. Enter the probability for each rating in columns 2, 4, 6, 8, 10, and 12 to the right. Note: this sum of probabilities must equal 1.0.
3. Form the entries on each row of col. 13; these are the components of the mean DEMs, according to equation (5.8).

$$13 = 1 \cdot 2 + 3 \cdot 4 + 5 \cdot 6 + 7 \cdot 8 + 9 \cdot 10 + 11 \cdot 12$$

4. At the bottom of col. 13, enter the product of all the entries in that column.
5. Form the entries on each row of col. 14; these are the components of the standard deviation of the DEMs, according to equation (5.11).

$$14 = \left[(1 \cdot 2 - 13)^2 + (3 \cdot 4 - 13)^2 + (5 \cdot 6 - 13)^2 + (7 \cdot 8 - 13)^2 + (9 \cdot 10 - 13)^2 + (11 \cdot 12 - 13)^2 \right]^{1/2}$$

6. At the bottom of col. 14, enter the total of the non-zero values on the rows above.
7. Divide each row of col. 14 by the same row of col. 13; enter these values in col. 15.
8. Square each entry in col. 15, add these squares, and enter *the square root of this sum at the bottom of col.* 15.
9. The data at the bottom of cols. 13, 14, and 15 will be used later.

Figure 5-1. Instructions for Calculating Mean and Standard Deviation of DEMs

Mode: Embed.
Coeffs: a= 2.8 , b= 1.20

CSCI #, i= 1		DEMs: Rating/Probability						Mean and Standard Deviation of DEMs		
Column Number		1 2	3 4	5 6	7 8	9 10	11 12	13	14	15
DEM #	& Name	VL / P	L / P	N / P	H / P	VH / P	XH / P	$\overline{M}_{ij}$	$\sigma_{M_{ij}}$	$\sigma_{M_{ij}} / \overline{M}_{ij}$
1	RELY	.75/	.88/	1 / 1	1.15/	1.40/	_ / _	1		
2	DATA	_ / _	.94/	1 / 1	1.08/	1.16/	_ / _	1		
3	CPLX	.70/0.05	.85/ 0.10	1 / 0.25	1.15/ 0.35	1.30/ 0.20	1.65/ 0.05	1.115	0.120	0.107
4	TIME	_ / _	_ / _	1 / 0.50	1.11/ 0.50	1.30/	1.65/	1.055	0.055	0.052
5	STOR	_ / _	_ / _	1 / 1	1.06/	1.21/	1.56/	1		
6	VIRT	_ / _	.87/	1 / 1	1.15/	1.30/	_ / _	1		
7	TURN	_ / _	.87/	1 / 1	1.07/	1.15/	_ / _	1		
8	ACAP	1.46/	1.19/ 0.10	1 / 0.30	.86/0.50	.71/ 0.10	_ / _	0.920	0.066	0.072
9	AEXP	1.29/	1.13/	1 / 1	.91/	.82/	_ / _	1		
10	PCAP	1.42/	1.17/	1 / 1	.86/	.70/	_ / _	1		
11	VEXP	1.21/	1.10/	1 / 1	.90/	_ /	_ / _	1		
12	LEXP	1.14/	1.07/	1 / 1	.95/	_ /	_ / _	1		
13	MODP	1.24/	1.10/	1 / 1	.91/	.82/	_ / _	1		
14	TOOL	1.24/	1.10/	1 / 1	.91/	.83/	_ / _	1		
15	SCED	1.23/	1.08/	1 / 1	1.04/	1.10/	_ / _	1		
16										
17										
18										
19										
20										
Bottom								$\overline{M}_i = 1.082$	$NZM_i = 3$	$\sigma_{M_i} / \overline{M}_i = 0.139$

Figure 5-2. Calculation of Mean and Standard Deviation of DEMs in CSCI 1

Mode: Embed.

Coeffs: a= 2.8 , b= 1.20

CSCI #, i= 2		DEMs: Rating/Probability						Mean and Standard Deviation of DEMs		
Column Number		1 2	3 4	5 6	7 8	9 10	11 12	13	14	15
DEM #	& Name	VL / P	L / P	N / P	H / P	VH / P	XH / P	$\bar{M}_{ij}$	$\sigma_{M_{ij}}$	$\sigma_{M_{ij}} / \bar{M}_{ij}$
1	RELY	.75/	.88/	1 /	1.15/	1.40/	_ / _	1		
2	DATA	_ / _	.94/	1 / 0.70	1.08/ 0.30	1.16/	_ / _	1.024	0.037	0.036
3	CPLX	.70/	.85/	1 /	1.15/	1.30/	1.65/	1		
4	TIME	_ / _	_ / _	1 /	1.11/	1.30/	1.65/	1		
5	STOR	_ / _	_ / _	1 /	1.06/ 0.60	1.21/ 0.40	1.56/	1.120	0.073	0.066
6	VIRT	_ / _	.87/	1 /	1.15/	1.30/	_ / _	1		
7	TURN	_ / _	.87/	1 /	1.07/	1.15/	_ / _	1		
8	ACAP	1.46/	1.19/	1 /	.86/	.71/	_ / _	1		
9	AEXP	1.29/	1.13/	1 /	.91/	.82/	_ / _	1		
10	PCAP	1.42/	1.17/ 0.50	1 / 0.50	.86/	.70/	_ / _	1.085	0.134	0.124
11	VEXP	1.21/	1.10/	1 /	.90/	_ /	_ / _	1		
12	LEXP	1.14/	1.07/	1 /	.95/	_ /	_ / _	1		
13	MODP	1.24/	1.10/	1 /	.91/	.82/	_ / _	1		
14	TOOL	1.24/	1.10/	1 /	.91/	.83/	_ / _	1		
15	SCED	1.23/	1.08/	1 /	1.04/	1.10/	_ / _	1		
16										
17										
18										
19										
20										
Bottom								$\bar{M}_i$ = 1.244	NZM_i = 3	$\sigma_{M_i} / \bar{M}_i$ =0.145

Figure 5-3. Calculation of Mean and Standard Deviation of DEMs in CSCI 2

Mode: Embed.

Coeffs: a= 2.8 , b= 1.20

CSCI #, i= 3		DEMs: Rating/Probability						Mean and Standard Deviation of DEMs		
Column Number		1 2	3 4	5 6	7 8	9 10	11 12	13	14	15
#	DEM & Name	VL / P	L / P	N / P	H / P	VH / P	XH / P	$\overline{M}_{ij}$	$\sigma_{M_{ij}}$	$\sigma_{M_{ij}} / \overline{M}_{ij}$
1	RELY	.75/	.88/	1 /	1.15/ 0.50	1.40/0.50	_ / _	1.275	0.125	0.098
2	DATA	_ / _	.94/	1 / 1	1.08/	1.16/	_ / _	1		
3	CPLX	.70/	.85/	1 / 1	1.15/	1.30/	1.65/	1		
4	TIME	_ / _	_ / _	1 / 1	1.11/	1.30/	1.65/	1		
5	STOR	_ / _	_ / _	1 / 1	1.06/	1.21/	1.56/	1		
6	VIRT	_ / _	.87/	1 / 1	1.15/	1.30/	_ / _	1		
7	TURN	_ / _	.87/	1 / 1	1.07/	1.15/	_ / _	1		
8	ACAP	1.46/	1.19/	1 / 1	.86/	.71/	_ / _	1		
9	AEXP	1.29/	1.13/ 0.25	1 / 0.50	.91/ 0.25	.82/	_ / _	1.010	0.078	0.078
10	PCAP	1.42/	1.17/	1 / 1	.86/	.70/	_ / _	1		
11	VEXP	1.21/	1.10/	1 / 1	.90/	_ /	_ / _	1		
12	LEXP	1.14/	1.07/ 0.20	1 / 0.70	.95/ 0.10	_ /	_ / _	1.009	0.034	0.034
13	MODP	1.24/	1.10/	1 / 1	.91/	.82/	_ / _	1		
14	TOOL	1.24/	1.10/	1 / 1	.91/	.83/	_ / _	1		
15	SCED	1.23/	1.08/	1 / 1	1.04/	1.10/	_ / _	1		
16										
17										
18										
19										
20										
Bottom								$\overline{M}_i$ = 1.300	NZM_i = 3	$\sigma_{M_i} / \overline{M}_i$ =0.130

Figure 5-4. Calculation of Mean and Standard Deviation of DEMs in CSCI 3

This form is used for calculating the mean and standard deviation of size for the several CSCIs and of effort for the entire system.

1. Enter the size range data for each CSCI on the appropriate row in columns 1, 2, and 3 if the size pdf is triangular. If the size pdf is uniform, leave col. 2 blank as an indication thereof.

2. For each CSCI, calculate the mean size according to (5.5) if the size pdf is uniform, and according to (5.7) if the size pdf is triangular. Enter the mean sizes in col. 4.

3. Sum the entries in col. 4; enter the sum at the bottom of col. 4.

4. Calculate the Nominal Effort according to $\bar{N} = a\bar{I}^b$ using the selected values of a and b, above, and the value of mean size from the bottom of col. 4, enter $\bar{N}$ at the bottom of col. 5.

5. Calculate each row of col. 5 as: 5 = 4 * (5-Bottom)/(4-Bottom).

6. On each row of col. 6, enter the value of M_i from the bottom of col. 13 of figures 5-2, 5-3, and 5-4, as appropriate. Then, in each row of col. 7, enter the product of the values in col. 5 and col. 6; enter the sum of all entries in this column at the bottom of this column. This sum is the point estimate of effort.

7. Calculate the standard deviation of size of each CSCI, using (5.9) if a uniform pdf, or using (5.10) if a triangular pdf has been selected. Divide these standard deviations by the mean size from the bottom of col. 4, and enter this normalized standard deviation on the corresponding row of Col. 8.

8. On each row of col. 9, calculate the value of the coefficients T_i, using (5.5).

Figure 5-5. Instructions for Computation of Mean and Standard Deviation of Effort

9. On each row of col. 10, enter the product of col. 9 with the square of the value in col. 8. Enter the sum of all entries in the space at the bottom of col. 10. Sum this with the value of the point estimate below 7; this sum is the estimated mean effort.

10. On each row of col. 11, calculate the coefficients S_i, according to (5.4).

11. On each row of col. 12, enter the product of col. 7 with col. 11. At the bottom of this column, enter the square root of the sum of the squares of all entries above. This is the standard deviation of effort due to the uncertainties of size.

12. On each row of col. 13, enter the normalized standard deviation of the product of the DEMs for the CSCI, from figures 5-2, 5-3, or 5-4, as appropriate for that CSCI.

13. On each row of col.14, enter the product of col. 7 with col. 13. Enter the square root of the sum of all entries in this column at the bottom; this is the standard deviation of effort due to the uncertainty of the DEMs.

14. Use the test criteria at the bottom of the form: enter the standard deviation of size from the bottom of col. 10, and that for the uncertaimties of the DEMs from the bottom of col. 14, and verify the second criterion.

15. The uncertainty of the COCOMO model is approximately 16% of the point estimate; enter 0.16 times the point estimate from the bottom of col. 7.

16. The standard deviation of the effort is the square root of the sum of the squares of the COCOMO model uncertainty, the uncertainty due to size at the bottom of col. 12, and the uncertainty of the DEMs at the bottom of col. 14. Enter this value below the bottom of col. 12 in the large space provided.

17. Calculate the probability cumulative distribution function, using the mean effort from the bottom of col. 9 and the standard deviation of the effort due to all sources, from the large space below col. 12.

Figure 5-5. (Concluded)

	Size Data			Mean Size & Effort							Std. Dev. of Eff. due to Size & DEMs			
COL #	1	2	3	4	5	6	7	8	9	10	11	12	13	14
CSCI #	L	M	H	$\bar{T}_i$	$\bar{N}_i$	$\bar{M}_i$	$\bar{D}_i$	$\sigma_{T_i}/\bar{T}$	T_i	ΔD_i	S_i	σ_{D_i}	$\sigma_{M_i}/\bar{M}_i$	σ_{D_i}
1	16	32	64	37.3	285	1.082	308	0.06621	136.1	0.415	1526.2	101.1	0.139	42.8
2	40	40	80	53.3	407	1.244	506	0.06256	173.4	0.678	1712.5	107.1	0.145	73.4
3	40	-	40	60.0	458	1.300	595	0.07662	186.2	1.093	1776.9	136.1	0.130	77.4
4														
5														
6														
7														
8														
9														
10														
Bottom			NZS= 3	$\bar{T}$=	$\bar{N}$=		$\bar{D}_0$=			ΔD= 2.2		σ_{D_S} = 200.5		σ_{D_M} =114.9
			NZM= 9	151	1150		1410	$\bar{D}$= 1412			$\sigma_{D_{Model}}$ = 225.5		$\sigma_{D_{Total}}$ = 322.9	
			NZT= 12											

Test criteria: 1. NZT= 12 > 2

2. Max. variance = $(136.1)^2 = 18523 < \dfrac{\left[(\sigma_{D_{size}})^2 + (\sigma_{D_{DEMs}})^2 \right]}{2} = \dfrac{[(200.5)^2 + (114.9)^2]}{2} = 26701$

Figure 5-6. Calculation of Mean and Standard Deviation of Effort

Table 5-3. Cumulative Probability vs. Effort

Standard Deviations, x	Effort, SM	Cumulative Probability, Percent
-1.28	1000	10.1
-1.12	1050	13.1
-0.97	1100	16.7
-0.81	1150	20.9
-0.66	1200	25.6
-0.50	1250	30.8
-0.35	1300	36.4
-0.19	1350	42.4
-0.04	1400	48.5
0.00	1412	50.0
0.27	1450	54.7
0.27	1500	60.7
0.43	1550	66.5
0.58	1600	72.0
0.74	1650	77.0
0.89	1700	81.4
1.05	1750	85.2
1.20	1800	88.5
1.36	1850	91.3
1.51	1900	93.5
1.67	1950	95.2
1.82	2000	96.6

Thus, for example, the probability that the project will require not more than 1700 SM is 81.4 percent while, by interpolation, it is 70 percent probable that the effort will not exceed 1580 SM.

Alternately, normal probability graph paper may be used. On this type of paper a normal CDF is a straight line. Enter the following pairs on the paper: $((\bar{D}-\sigma_D)$, 15.9%), ($\bar{D}$, 50%), and $((\bar{D}+\sigma_D)$, 84.1%); the three points will lie on a straight line from which table 5-3 may be formed directly; see figure 5-7. As a further alternate, an excellent algorithm for the normal CDF is given in [5].

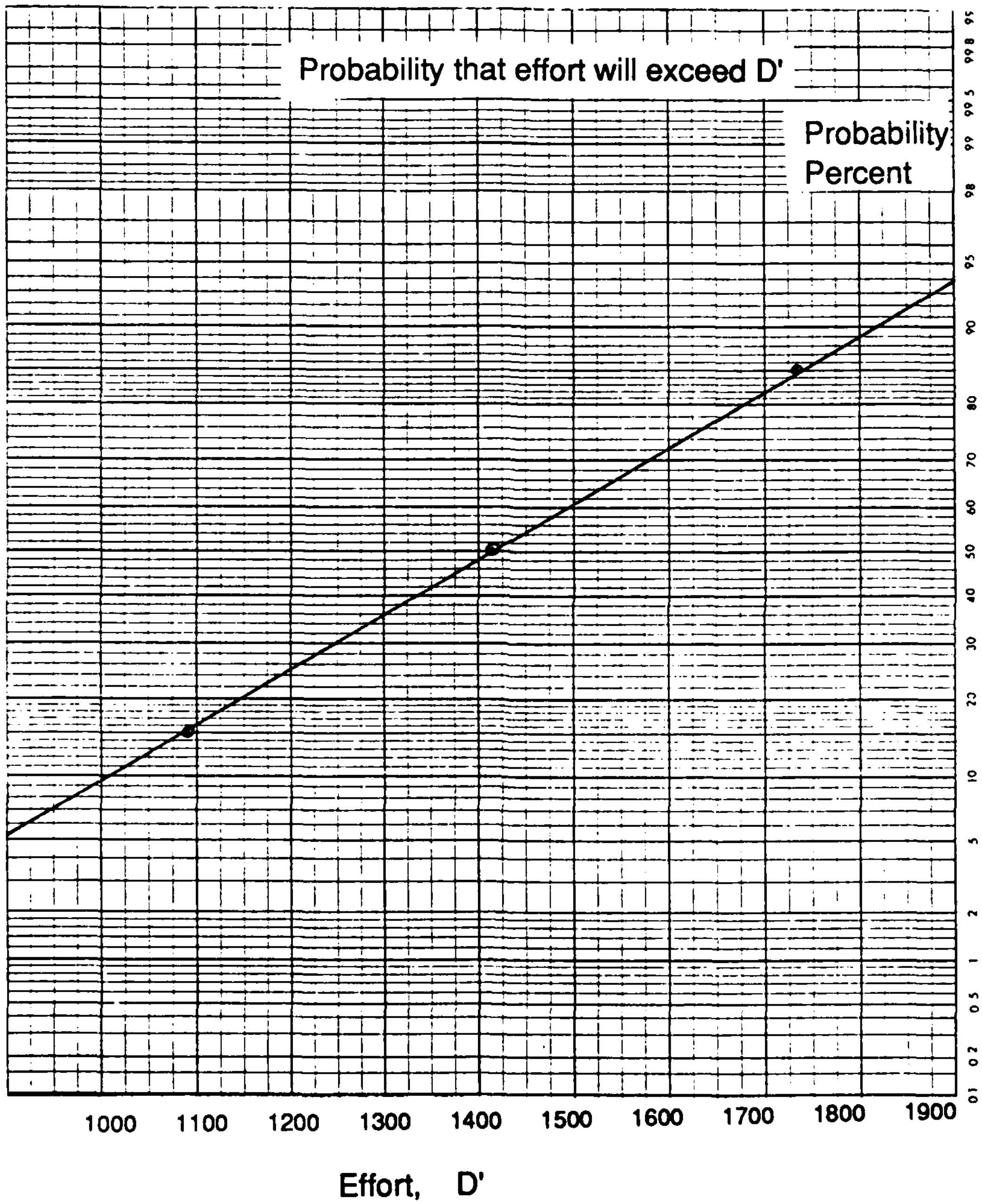

Figure 5-7. Cumulative Distribution Function of Effort

LIST OF REFERENCES

1. Boehm, Barry W., 1981, Software Engineering Economics, Englewood Cliffs, NJ: Prentice-Hall, Inc.

2. Gibra, Isaac N., 1973, Probability and Statistical Inference for Scientists and Engineers, Englewood Cliffs, NJ: Prentice-Hall, Inc.

3. Saaty, Thomas L., 1959, Mathematical Methods of Operations Research, York, PA: Maple Press Co.

4. Garvey, Paul R., November 1987, The Effect of Software Size Uncertainty on Effort Estimates Generated by αI^{β} Software Resource Models, ESD-TR-87-169, AD A189 209, Electronic Systems Division, AFSC, Hanscom Air Force Base, MA.

5. Abramowitz, M., and I. A. Stegun, (eds.), 1965, Handbook of Mathematical Functions, NY: Dover Publications.

APPENDIX A

This appendix gathers the various computation forms which were used in section 5, together with the various instruction sheets. This appendix therefore contains all the material necessary to perform a complete software effort uncertainty analysis, or to form spreadsheets for the computation.

This form is used for calculating the mean and the standard deviation of the product of the DEMs in each CSCI for Intermediate COCOMO. *The outputs are the mean* product of the DEMs and the normalized standard deviation of the product of the DEMs. The mean appears at the bottom of column 13, while the normalized standard deviation appears at the bottom of col. 15, and the number of uncertain items is given at the bottom of col. 14. Detailed instructions follow.

1. At upper-right corner, enter the Mode (Organic, Semi-det...., Embedded), and the COCOMO coefficients a and b for that mode. Inside form, at upper-left corner, enter the CSCI number (i=).
2. The ratings for each attribute for Intermediate COCOMO are given in columns 1, 3, 5, 7, 9, 11. Enter the probability for each rating in columns 2, 4, 6, 8, 10, *and* 12 *to the right. Note: this sum of probabilities must* equal 1.0.
3. Form the entries on each row of col. 13; these are the components of the mean DEMs, according to equation (5.8).

$$13 = 1 \cdot 2 + 3 \cdot 4 + 5 \cdot 6 + 7 \cdot 8 + 9 \cdot 10 + 11 \cdot 12$$

4. At the bottom of col. 13, enter the product of all the entries in that column.
5. Form the entries on each row of col. 14; these are the components of the standard deviation of the DEMs, according to equation (5.11).

$$14 = \left[(1 \cdot 2 - 13)^2 + (3 \cdot 4 - 13)^2 + (5 \cdot 6 - 13)^2 + (7 \cdot 8 - 13)^2 + (9 \cdot 10 - 13)^2 + (11 \cdot 12 - 13)^2 \right]^{1/2}$$

6. At the bottom of col. 14, enter the total of the non-zero values on the rows above.
7. Divide each row of col. 14 by the same row of col. 13; enter these values in col. 15.
8. Square each entry in col. 15, add these squares, and enter the square root of this sum at the bottom of col. 15.
9. The data at the bottom of cols. 13, 14, and 15 will be used later.

Figure A-1. Instructions for Calculating Mean and Standard Deviation of DEMs

Mode:

Coeffs: a= , b=

CSCI #, i=		DEMs:			Rating/Probability			Mean and Standard Deviation of DEMs		
Column Number		1 2	3 4	5 6	7 8	9 10	11 12	13	14	15
DEM #	& Name	VL / P	L / P	N / P	H / P	VH / P	XH / P	$\overline{M_{ij}}$	σM_{ij}	$\sigma M_{ij} / \overline{M_{ij}}$
1	RELY	.75/	.88/	1 /	1.15/	1.40/	_ / _			
2	DATA	_ / _	.94/	1 /	1.08/	1.16/	_ / _			
3	CPLX	.70/	.85/	1 /	1.15/	1.30/	1.65/			
4	TIME	_ / _	_ / _	1 /	1.11/	1.30/	1.65/			
5	STOR	_ / _	_ / _	1 /	1.06/	1.21/	1.56/			
6	VIRT	_ / _	.87/	1 /	1.15/	1.30/	_ / _			
7	TURN	_ / _	.87/	1 /	1.07/	1.15/	_ / _			
8	ACAP	1.46/	1.19/	1 /	.86/	.71/	_ / _			
9	AEXP	1.29/	1.13/	1 /	.91/	.82/	_ / _			
10	PCAP	1.42/	1.17/	1 /	.86/	.70/	_ / _			
11	VEXP	1.21/	1.10/	1 /	.90/	_/	_ / _			
12	LEXP	1.14/	1.07/	1 /	.95/	_/	_ / _			
13	MODP	1.24/	1.10/	1 /	.91/	.82/	_ / _			
14	TOOL	1.24/	1.10/	1 /	.91/	.83/	_ / _			
15	SCED	1.23/	1.08/	1 /	1.04/	1.10/	_ / _			
16										
17										
18										
19										
20										
Bottom								$\bar{M}_i$ =	NZM_i =	$\sigma M_i / \bar{M}_i$ =

Figure A-2. Form for Computation of Mean and Standard Deviation of DEMs

This form is used for calculating the mean and standard deviation of size for the several CSCIs and of effort for the entire system.

1. Enter the size range data for each CSCI on the appropriate row in columns 1, 2, and 3 if the size pdf is triangular. If the size pdf is uniform, leave col. 2 blank as an indication thereof.

2. For each CSCI, calculate the mean size according to (5.5) if the size pdf is uniform, and according to (5.7) if the size pdf is triangular. Enter the mean sizes in col. 4.

3. Sum the entries in col. 4; enter the sum at the bottom of col. 4.

4. Calculate the Nominal Effort according to $\bar{N}=a\bar{I}^b$ using the selected values of a and b, above, and the value of mean size from the bottom of col. 4, enter $\bar{N}$ at the bottom of col. 5.

5. Calculate each row of col. 5 as: 5 = 4 * (5-Bottom)/(4-Bottom).

6. On each row of col. 6, enter the value of M_i from the bottom of col. 13 of figures A-2, as appropriate. Then, in each row of col. 7, enter the product of the values in col. 5 and col. 6; enter the sum of all entries in this column at the bottom of this column. This sum is the point estimate of effort.

7. Calculate the standard deviation of size of each CSCI, using (5.9) if a uniform pdf, or using (5.10) if a triangular pdf has been selected. Divide these standard deviations by the mean size from the bottom of col. 4, and enter this normalized standard deviation on the corresponding row of Col. 8.

8. On each row of col. 9, calculate the value of the coefficients T_i, using (5.5).

Figure A-3. Instructions for Computation of Mean and Standard Deviation of Effort

9. On each row of col. 10, enter the product of col. 9 with the square of the value in col. 8. Enter the sum of all entries in the space at the bottom of col. 10. Sum this with the value of the point estimate below 7; this sum is the estimated mean effort.

10. On each row of col. 11, calculate the coefficients S_i, according to (5.4).

11. On each row of col. 12, *enter the product of col. 7 with col. 11. At the bottom of this column, enter* the square root of the sum of the squares of all entries above. This is the standard deviation of effort due to the uncertainties of size.

12. On each row of col. 13, enter the normalized standard deviation of the product of the DEMs for the CSCI, from figures A-2, as appropriate for that CSCI.

13. On each row of col.14, *enter the product of col. 7 with col. 13. Enter the square root of the sum of all* entries in this column at the bottom; this is the standard deviation of effort due to the uncertainty of the DEMs.

14. Use the test criteria at the bottom of the form: enter the standard deviation of size from the bottom of col. 10, and that for the uncertainties of the DEMs from the bottom of col. 14, and verify the second criterion.

15. The uncertainty of the COCOMO model is approximately 16% of the point estimate; enter 0.16 times the point estimate from the bottom of col. 7.

16. The standard deviation of the effort is the square root of the sum of the squares of the COCOMO model uncertainty, the uncertainty due to size at the bottom of col. 12, and the uncertainty of the DEMs at the bottom of col. 14. Enter this value below the bottom of col. 12 in the large space provided.

17. Calculate the probability cumulative distribution function, using the mean effort from the bottom of col. 9 and the standard deviation of the effort due to all sources, from the large space below col. 12.

Figure A-3. (Concluded)

	Size Data			Mean Size & Effort							Std. Dev. of Eff. due to Size & DEMs			
COL #	1	2	3	4	5	6	7	8	9	10	11	12	13	14
CSCI #	L	M	H	$\bar{T}_i$	$\bar{N}_i$	$\bar{M}_i$	$\bar{D}_i$	$\sigma_{Ti}/\bar{T}$	T_i	ΔD_i	S_i	σ_{Di}	$\sigma_{Mi}/\bar{M}_i$	σ_{Di}
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
Bottom			NZS=	$\bar{T}$=	$\bar{N}$=		D_0 =			Δ D=		σ_{D_S} =		σ_{D_M} =
			NZM					$\bar{D}$=			$\sigma_{D_{Model}}$ =		$\sigma_{D_{Total}}$ =	
			NZT=											

Test criteria: 1. NZT= > 2

2. Max. variance = $(\quad)^2 =$ $< \dfrac{\left[(\sigma_{D_{size}})^2 + (\sigma_{D_{DEMs}})^2\right]}{2} = \dfrac{[(\quad)^2 + (\quad)^2]}{2} =$

Figure A-4. Form for Computation of Mean and Standard Deviation of Effort

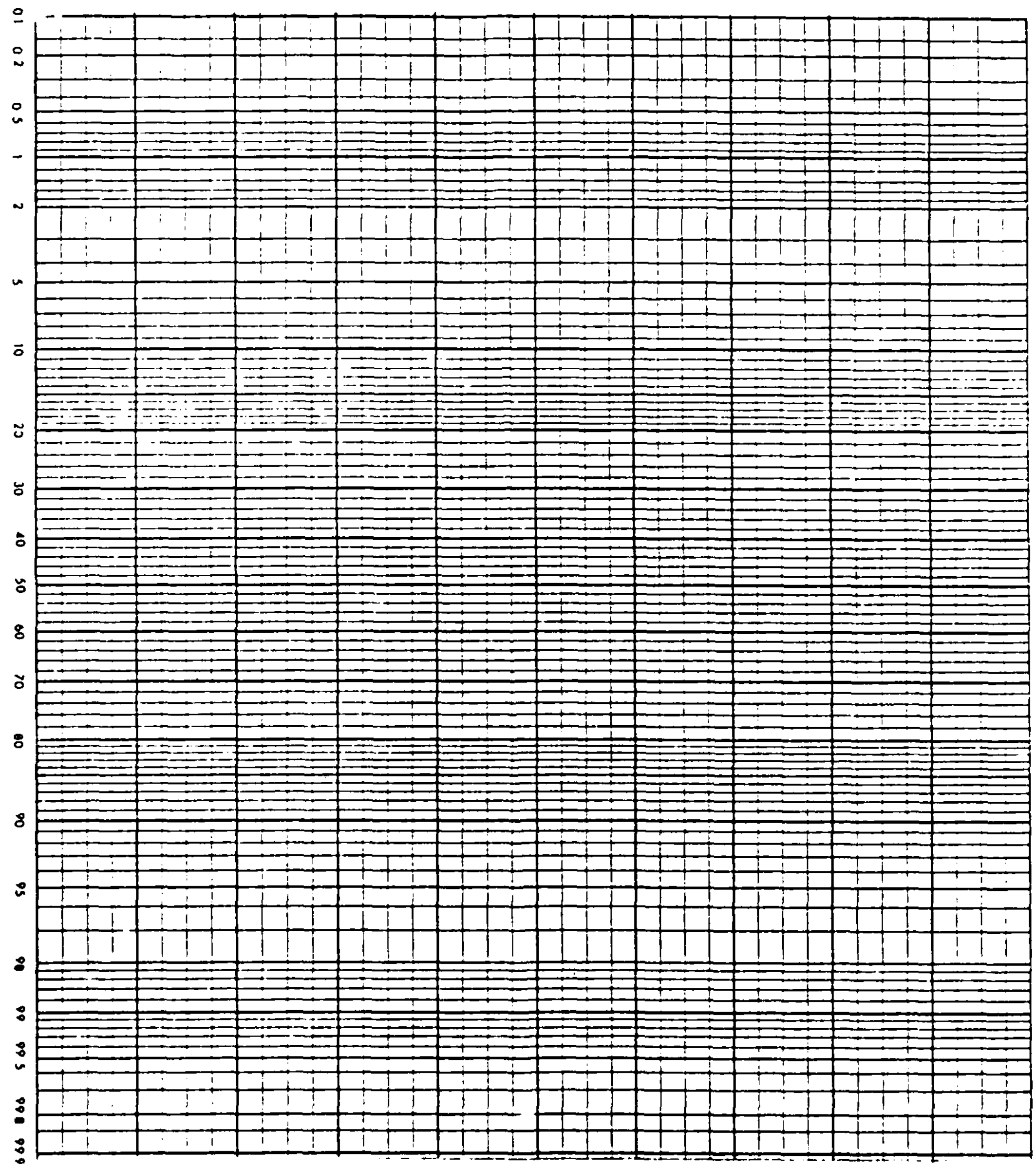

Figure A-5. Normal Cumulative Probability Paper